Must we settle for less rigorous evaluations in large area-based crime prevention programs? Lessons from a Campbell review of focused deterrence

Anthony A. Braga · David L. Weisburd

Abstract

Objectives Evaluations from a recent Campbell systematic review of focused deterrence programs are critically reviewed to determine whether more rigorous evaluations are possible given methodological challenges such as developing appropriate units of analysis, generalizing findings beyond study sites, and controlling for the contamination of available comparison groups.

Methods We synthesize the available evaluation literature on focused deterrence programs completed before and after the publication of the Campbell review to assess opportunities to conduct randomized controlled trials and stronger quasi-experimental evaluations.

Results We find that focused deterrence strategies are amenable to more rigorous evaluation methodologies such as block randomized place-based trials, multisite cluster randomized trials, and quasi-experimental evaluations that employ advanced statistical matching techniques.

Conclusions Focused deterrence programs can, and should, be subjected to more rigorous tests that generate more robust evidence on program impacts and provide further insight into the crime control mechanisms at work in these programs. More generally, our review supports the idea that program evaluators do not have to "settle for less" methodological rigor when testing large area-based crime prevention programs.

A. A. Braga
Rutgers University, Newark, NJ, USA

A. A. Braga (✉)
Harvard University, Cambridge, MA, USA
e-mail: Anthony_Braga@harvard.edu

D. L. Weisburd
Hebrew University, Jerusalem, Israel

D. L. Weisburd
George Mason University, Fairfax, VA, USA

🖄 Springer

Introduction

Focused deterrence strategies are a relatively new addition to a growing portfolio of evidence-based crime prevention practices available to policy makers and practitioners. Briefly, focused deterrence strategies seek to change offender behavior by understanding underlying crime-producing dynamics and conditions that sustain recurring crime problems and by implementing a blended strategy of law enforcement, community mobilization, and social service actions (Kennedy 1997, 2008). Direct communications of increased enforcement risks and the availability of social service assistance to target groups and individuals is a defining characteristic of focused deterrence programs. In response to conflicting reports on the crime control efficacy of these new crime prevention strategies (see, e.g., Braga et al. 2001; Rosenfeld et al. 2005; Wellford et al. 2005), the United Kingdom's National Policing Improvement Agency (NPIA) provided funds to support a Campbell Collaboration systematic review of the available evaluation evidence on the crime control efficacy of focused deterrence strategies. The Campbell review found that focused deterrence strategies were associated with significant reductions in targeted crime problems but strongly recommended that the available evaluation evidence needed to be strengthened (Braga and Weisburd 2012).

Some observers, however, trumpeted the Campbell review as definitive evidence that focused deterrence strategies "work" in controlling crime. For instance, the Center for Crime Prevention and Control at John Jay College of Criminal Justice described the Campbell review as "the gold standard in evaluating social science interventions" and highlighted the "strong empirical evidence" in support of focused deterrence strategies.[1] While Campbell reviews generally only include more rigorous controlled evaluations, the strength of the review is rooted in the quality of available evidence. While Braga and Weisburd (2012) did conclude the available evidence was highly supportive of crime reduction impacts, they also noted that existing focused deterrence program evaluations were completely comprised of quasi-experimental tests, many of which were weaker designs that used non-equivalent comparisons. They expressed concern over the lack of randomized controlled trials and called for more rigorous evaluations of focused deterrence strategies. Unfortunately, to date, that call has not been answered.

A more careful interpretation of the Campbell review would be that there is some promising evidence that focused deterrence strategies do indeed generate significant crime reduction impacts. However, these strategies need to be subjected to more rigorous tests that generate more robust evidence on program impacts and, as suggested by Braga and Weisburd (2012), provide further insight into the crime control mechanisms at work in these programs. Of course, a key question is whether such evaluations can actually be implemented. A number of scholars suggest that we will have to "settle for less" in the evaluation of many crime prevention programs, like focused deterrence strategies (Eck 2002; Knutsson 2009; Pawson and Tilley 1997; Tilley 2009). For instance, these scholars argue that many targeted crime problems are unique and it is

---

[1] http://johnjayresearch.org/ccpc/campbell-collaboration/

nearly impossible to find a "counterfactual" control group for them. In turn, many focused deterrence strategies, as we describe below, are implemented to change violent behavior among social networks of offenders that often span large areas, and thus it is difficult to imagine large field experiments with many sites.

Our own view is that evaluations of focused deterrence can use more rigorous designs. This will not be easy but it is, as we detail below, possible. In this paper, we review the designs used in evaluating focused deterrence programs. We then discuss why scholars might argue that evaluators have to "settle for less" in developing solid empirical evidence in this area. We focus on two possibilities that would allow more rigorous evaluations. The first is to develop stronger quasi-experimental evaluations. The second is to conduct some much needed randomized controlled trials of focused deterrence programs. As we outline below, we think that randomized field experiments can and should be used in this area so that we can develop stronger evidence about how to prevent gang and group-involved violence, reduce recidivism by high-rate offenders, and control violent drug markets. We also think that we can draw broader lessons for improving crime prevention evaluations which are focused in large areas and have to date been seen as demanding less rigorous quasi-experimental models of evaluation.

Key findings of the campbell systematic review on focused deterrence programs

Focused deterrence strategies honor core deterrence ideas, such as increasing risks faced by offenders, while finding new and creative ways of deploying traditional and non-traditional law enforcement tools to do so, such as directly communicating incentives and disincentives to targeted offenders (Kennedy 1997, 2008). The focused deterrence approach is also consistent with recent theorizing about police innovation, which suggests that approaches that seek to both create more focus in application of crime prevention programs and that expand the tools of policing are likely to be most successful (Weisburd and Eck 2004). The available scientific evidence on the crime reduction value of focused deterrence strategies had been previously characterized as "promising", but "descriptive rather than evaluative" (Skogan and Frydl 2004: 241), and as "limited" but "still evolving" (Wellford et al. 2005: 10), by the U.S. National Research Council's Committee to Review Research on Police Policy and Practices and Committee to Improve Research Information and Data on Firearms, respectively.

Braga and Weisburd (2012) identified ten focused deterrence evaluations in their Campbell review; eight of which were completed after the National Research Council reports were published. A better-developed base of scientific evidence now exists to assess whether crime prevention impacts are associated with this approach. The ten studies included in the Campbell review included:

1. Operation Ceasefire in Boston (Braga et al. 2001)
2. Operation Ceasefire in Los Angeles (Tita et al. 2004)
3. Indianapolis Violence Reduction Partnership (McGarrell et al. 2006)
4. Project Safe Neighborhoods in Chicago (Papachristos et al. 2007)
5. Operation Peacekeeper in Stockton (Braga 2008)
6. Project Safe Neighborhoods in Lowell (Braga et al. 2008b)
7. Drug Market Intervention in Nashville (Corsaro and McGarrell 2009)

8. Drug Market Intervention in Rockford (Corsaro et al. 2010)
9. Cincinnati Initiative to Reduce Violence (Engel et al. 2010)
10. Operation Ceasefire in Newark (Boyle et al. 2010)

Six studies evaluated the crime reduction effects of focused deterrence pulling levers strategies on serious violence generated by street gangs or criminally-active street groups (Boston, Cincinnati, Indianapolis, Los Angeles, Lowell, and Stockton). Drawing on the Boston experience (Kennedy et al. 1996), these group-based violence reduction strategies join together criminal justice agencies, social service organizations, and community members to directly engage with violent groups and clearly communicate credible moral and law enforcement messages against violence, make genuine offers of help for those who want it, and launch strategic enforcement campaigns against those who continue their violent behavior (Kennedy 2008).

Two studies evaluated strategies focused on reducing crime driven by street-level drug markets (Nashville and Rockford) and are generally called "Drug Market Intervention" (DMI)-focused deterrence strategies. DMI-focused deterrence strategies identify street-level dealers, immediately apprehend violent drug offenders, and suspend criminal cases for non-violent dealers (Kennedy 2008). DMI strategies then bring together non-violent drug dealers, their families, law enforcement and criminal justice officials, service providers, and community leaders for a meeting that makes clear the dealing has to stop, the community cares for the offenders but reject their conduct, help is available, and renewed dealing will result in the activation of the existing case (Kennedy and Wong 2009). Two studies evaluated crime reduction strategies that were focused on individuals (Chicago and Newark). In general, these strategies address the most dangerous offenders with a wide range of legal tools, put offenders on formal prior notice that a "next offense" will bring extraordinary legal attention, and focus community "moral voices" on such offenders to set a clear standard that violence is unacceptable (Kennedy 2008).

Braga and Weisburd (2012) raised some modest concerns over construct validity in the focused deterrence evaluations that were reviewed. While the evaluations were supportive of deterrence principles, they noted that it was difficult to know whether observed reductions represented a true deterrent impact. A growing number of scholars suggest that there seem to be additional crime control mechanisms at work in these strategies beyond straight-up deterrence (Braga 2012; Corsaro et al. 2012; Papachristos et al. 2007). Other prevention frameworks, such as community social control and procedural fairness, might help explain the observed impacts of focused deterrence programs on crime. In addition to advocating for more rigorous evaluation designs, Braga and Weisburd (2012) recommended that the next wave of research on focused deterrence strategies needs to develop further knowledge *why* these strategies seem to work.

Evaluation designs

All ten eligible studies used quasi-experimental designs to analyze the impact of pulling levers-focused deterrence strategies on crime (Braga and Weisburd 2012). Seven evaluations used quasi-experimental designs with non-equivalent comparison groups (Boston, Cincinnati, Indianapolis, Lowell, Nashville, Rockford, and Stockton). Two

evaluations used quasi-experimental designs with comparison groups created through matching techniques (Chicago and Newark). One evaluation used a quasi-experimental design that included both non-equivalent comparison groups and comparison groups created through matching techniques (Los Angeles).

Five studies evaluated the crime reduction effects of focused deterrence strategies by comparing trends in key outcome variables in a targeted geographic area (identified as a neighborhood, policing district, or well-defined zone) to trends in key outcome variables in comparison areas. The Chicago study used propensity score-matching techniques[2] to identify similar comparison policing districts to compare against the targeted policing districts. The Los Angeles study used two non-equivalent comparisons (the target area relative to the remainder of the larger neighborhood, and the targeted neighborhood relative to the surrounding larger geographic community area). The Los Angeles study also used propensity score-matching techniques to identify similar census block groups to compare against the census block groups that comprised the targeted area. The Newark evaluation used crime mapping technology and simple matching techniques to identify a comparison gun hot spot area that was similar to the targeted Operation Ceasefire zone in terms of gunshot wounding incidents, geographic size, and socio-demographic characteristics. The Nashville and Rockford studies compared crime trends in targeted neighborhoods relative to crime trends in the surrounding County and city areas, respectively.

Five studies evaluated the crime reduction effects of citywide pulling levers interventions. The Boston, Indianapolis, Lowell, and Stockton quasi-experimental designs compared citywide trends in key outcomes to citywide trends in key outcomes in sets of non-equivalent cities that did not experience a pulling levers intervention during the study time period. The Cincinnati evaluation compared citywide trends in homicides involving members of criminally-active groups targeted by the pulling levers intervention relative to trends in homicides that did not involve members of criminally active groups.

Using the Maryland Scientific Methods Scale (Sherman et al. 1997) as a standard, the Boston, Cincinnati, Indianapolis, Nashville, Rockford, and Stockton studies would be considered "Level 3" evaluations and also regarded as the minimum design that is adequate for drawing conclusions about program effectiveness. This design rules out many threats to internal validity such as history, maturation/trends, instrumentation, testing, and mortality. However, as Farrington et al. (2006) observe, the main problems of Level 3 evaluations center on selection effects and regression to the mean due to the non-equivalence of treatment and control conditions. The Chicago, Los Angeles, and Newark studies would be considered "Level 4" evaluations as they measured outcomes before and after the program in multiple treatment and control condition units. These types of designs have better statistical control of extraneous influences on the outcome and, relative to lower level evaluations, more adequately deal with selection and regression threats.
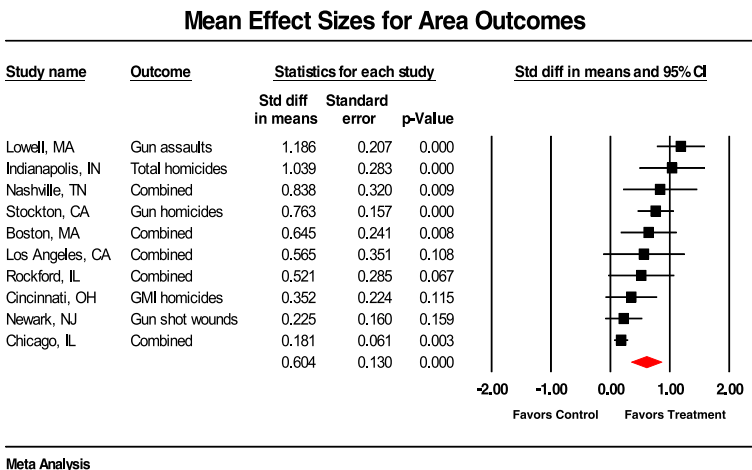
---

[2] Propensity score-matching techniques attempt to create equivalent treatment and comparison groups by summarizing relevant pre-treatment characteristics of each subject into a single-index variable (the propensity score) and then matching subjects in the untreated comparison pool to subjects in the treatment group based on values of the single-index variable (Rosenbaum and Rubin 1983, 1985).

Crime reduction impacts

Nine of the ten pulling levers-focused deterrence evaluations concluded that these programs generated significant crime control benefits (Braga and Weisburd 2012). While the authors did report a small but positive reduction in gunshot wound incidents, only the evaluation of Newark's Operation Ceasefire did not report any discernible crime prevention benefits generated by the violence reduction strategy. Evaluations of focused deterrence strategies targeting gangs and criminally active groups reported large statistically significant reductions in violent crime. These results included: a 63 % reduction in youth homicides in Boston (Braga et al. 2001), a 44 % reduction in gun assault incidents in Lowell (Braga et al. 2008b), a 42 % reduction in gun homicides in Stockton (Braga 2008), a 35 % reduction in homicides of criminally active group members in Cincinnati (Engel et al. 2010), a 34 % reduction in total homicides in Indianapolis (McGarrell et al. 2006), and noteworthy short-term reductions in violent crime in Los Angeles (Tita et al. 2004).

The two DMI evaluations also reported statistically significant crime reductions. In Nashville, the drug market intervention generated a 55 % reduction in illegal drug possession incidents (Corsaro and McGarrell 2009). In Rockford, the drug market intervention generated a 22 % reduction in non-violent offenses (Corsaro et al. 2010). While Newark's strategy did not generate statistically significant crime control gains when high-rate offenders were targeted, the Chicago PSN intervention, the other program focused on individuals, was associated with a 37 % reduction in homicide (Papachristos et al. 2007).

Following Campbell Collaboration protocols, Braga and Weisburd (2012) used meta-analyses of program effects to determine the size and direction of the effects and weighting effect sizes based on the variance of the effect size and the study sample size (Lipsey and Wilson 2001). The forest plots in Fig. 1 show the standardized difference in means between the treatment and control or comparison conditions (effect size) with a 95 % confidence interval plotted around them for all eligible studies. Points

Mean Effect Sizes for Area Outcomes

Study name	Outcome	Std diff in means	Standard error	p-Value	Std diff in means and 95% CI
Lowell, MA	Gun assaults	1.186	0.207	0.000	
Indianapolis, IN	Total homicides	1.039	0.283	0.000	
Nashville, TN	Combined	0.838	0.320	0.009	
Stockton, CA	Gun homicides	0.763	0.157	0.000	
Boston, MA	Combined	0.645	0.241	0.008	
Los Angeles, CA	Combined	0.565	0.351	0.108	
Rockford, IL	Combined	0.521	0.285	0.067	
Cincinnati, OH	GMI homicides	0.352	0.224	0.115	
Newark, NJ	Gun shot wounds	0.225	0.160	0.159	
Chicago, IL	Combined	0.181	0.061	0.003	
		0.604	0.130	0.000	

Meta Analysis

Fig. 1 Mean effect sizes for area outcomes in eligible focused deterrence evaluations. Source: Braga and Weisburd (2012: 48)

plotted to the right of 0 indicate a treatment effect; in this case, the study showed a reduction in crime or disorder. Points to the left of 0 indicate an effect where control conditions improved relative to treatment conditions. The meta-analysis of effect sizes suggests a strongly significant effect in favor of pulling levers-focused deterrence strategies. The overall effect size for these studies is 0.604. This is above Cohen's (1988) standard for a medium effect of 0.50 and below that of a large effect at 0.80. Nonetheless, the overall effect size is relatively large compared to assessments of interventions in crime and justice work more generally.

Given the important distinction in methodological quality between the non-equivalent quasi-experiments and the quasi-experiments that used matching techniques to identify comparison groups, Braga and Weisburd (2012) also examined research design as a moderator variable. The non-equivalent quasi-experimental designs were associated with a much larger within-group effect size (0.766, $p<0.05$) relative to the matched quasi-experimental designs (0.196, $p<0.05$). While the biases in quasi-experimental research are not clear (e.g., Campbell and Boruch 1975; Wilkinson and Task Force on Statistical Inference 1999), recent reviews in crime and justice suggest that weaker research designs often lead to more positive outcomes (e.g., see Weisburd et al. 2001; Welsh et al. 2011).

Should evaluators "settle" for weaker evaluations? The case of Boston ceasefire

Given the high profile of the seminal Boston experience during the mid- to late 1990s, the Braga et al. (2001) Operation Ceasefire evaluation has been reviewed by a number of researchers, and the relationship between the implementation of Ceasefire and the trajectory youth homicide in Boston during the 1990s has been closely scrutinized. Fagan (2002) suggested that some of the decrease in homicide may have occurred without the Ceasefire intervention in place as violence was decreasing in most major U.S. cities. In support of this perspective, Fagan (2002) presented a simple time-series graph on youth gun homicide in Boston and in other Massachusetts cities that suggested a general downward trend in gun violence may have existed before Ceasefire was implemented. Using growth-curve analysis to examine predicted homicide trend data for the 95 largest U.S. cities during the 1990s, (Rosenfeld et al. 2005) found some evidence of a sharper youth homicide drop in Boston than elsewhere, but suggest that the small number of youth homicide incidents precludes strong conclusions about program effectiveness based on their statistical models (see Berk 2005a for a critique of this evaluation).

Other reviewers, however, have been more supportive of a program effect in their reviews of the Ceasefire impact evaluation (e.g., Cook and Ludwig 2006). Ludwig (2005) suggested that Ceasefire was associated with a large drop in youth homicide but, given the complexities of analyzing city-level homicide trend data, there remained some uncertainty about the extent of Ceasefire's effect on youth violence in Boston. Morgan and Winship (2007) review of the Ceasefire evaluation concluded that the analysis was a "very high-quality example" of how to conduct an interrupted time series analysis of program impact and further noted that "they offer four types of supplemental analysis… which can be used to strengthen the warrant for causal assertion" (p. 252).

The National Academies' Panel on Improving Information and Data on Firearms (Wellford et al. 2005) concluded that the Ceasefire evaluation was compelling in associating the intervention with the subsequent decline in youth homicide. However, the Panel also suggested that many complex factors affect youth homicide trends and that it was difficult to specify the exact relationship between the Ceasefire intervention and subsequent changes in youth offending behaviors. The Panel further observed that the Ceasefire evaluation examined aggregate citywide data and did not provide any empirical evidence that treated gangs modified their violent behaviors after being exposed to the intervention.

Much of this uncertainty is due to the weak quasi-experimental design used to evaluate the 1990s Ceasefire focused deterrence program. And, unfortunately, the identified shortcomings of this well-known evaluation still contribute to contemporary debates over the crime control efficacy of focused deterrence strategies. For instance, a recent *Wall Street Journal* article on focused deterrence noted the scholarly debate over whether Boston's youth homicide decline during the 1990s could be attributed to the implementation of Ceasefire (Harless 2013). It is important to note here that more rigorous designs, including a randomized controlled trial, were initially considered for the original Boston Ceasefire evaluation (see Braga 2013). However, the Ceasefire working group, which included the Harvard research team, decided that more rigorous designs were not possible due to the group's strong desire to halt gang violence wherever and whenever it presented itself in the city. Moreover, the Ceasefire working group was also seeking to develop a new way of controlling outbreaks of gang violence and was concerned that a restrictive design may have impeded an innovative approach that attempted to modify the behavior of a very small and well-connected social network through a creative application of deterrence principles. As such, the implementation of Ceasefire proceeded with little further a priori attention given to evaluation design issues.

The specific nature of crime problems and the external validity of study findings

As we noted at the outset, a number of scholars have argued that it is unrealistic and often inappropriate to try to develop randomized experiments in areas like this, or even strong quasi-experimental studies. Much of this criticism has come from scholars who are as equally interested in developing policy recommendations as those who suggest a hierarchy of evidence. For example, situational crime prevention (e.g., Clarke 1997; Guerette 2009) and problem-oriented policing scholars (Eck 2002; Knutsson 2009) have argued that each "problem" is unique, and, accordingly, it is virtually impossible to develop even strong quasi-experiments for evaluating problem-oriented strategies such as those programs evaluated in focused deterrence studies. This position suggests that program evaluators should be comfortable making inferences about impacts based on one-group-only analyses of time series data.

And even if you can find specific comparisons, such studies are seen as likely to be very limited in their policy relevance (Pawson and Tilley 1997; Tilley 2009). In this case, scholars argue that randomized experiments are often so narrow in their focus, and in the samples they can define, that they fail in making generalizations to the population of problems about which we would like to make decisions. This is generally referred to as the problem of the external validity of study findings. External validity

gauges the extent to which the findings of a study can be generalized to the population of interest. Accordingly, external validity measures whether the results of a study have meaning for the "real" world of crime and justice that we are concerned with (Cook and Campbell 1979). A study can have very high internal validity but be relevant only to a very limited number of contexts or problems. Clearly, strongly designed studies should be capable of being generalized widely. Inferences about cause-effect relationships based on a specific scientific study are said to possess external validity if they may be generalized from the unique and idiosyncratic experimental settings, procedures, and participants to other populations and conditions.

Randomized experiments, in particular, are suggested to have lower external validity than other types of studies (Clarke and Cornish 1972; Pawson and Tilley 1997). This argument is often made when comparing observational non-experimental studies with randomized field trials (e.g., Sampson 2010). Randomized experiments are still not widely accepted by practitioners in criminal justice, and often require significant interventions in the daily routines of criminal justice agencies to be implemented successfully. This often means that only the most progressive criminal justice agencies are willing to be involved in randomized experiments. One consequence of this is that experiments are conducted in relatively special environments, ones which are willing and able to participate in a randomized study. We know of no study to date that has actually shown a relationship between study design and external validity, and we suspect that differences are not substantial (see also Weisburd 2010). The problem of external validity should be kept in mind in reviewing all study findings regardless of the design used. The difficulties associated with generalizing over subjects, settings, times, interventions, and outcomes are not unique to randomized experiments (Berk 2005b).

All applied crime prevention program evaluations can suffer from external validity concerns regardless of the degree of internal rigor in the evaluation research design. For instance, problem-oriented policing is primarily an analytic approach to crime prevention that requires customizing interventions to highly localized crime and disorder problems (Goldstein 1990). What works in preventing a street robbery problem in the public areas of Harvard Square in Cambridge, Massachusetts, might not work when applied to repeated robberies occurring in the London Underground subway system (Braga 2010). Appropriate interventions need to be applied in both contexts that are based on careful analysis of the conditions that create the compelling criminal opportunities. Neither the evaluation findings of a carefully constructed single group (no control group) before–after design nor the findings of a randomized controlled experiment will travel perfectly across these settings. Problem-oriented policing evaluations of many forms provide valuable guidance to police officers struggling with real-world problems. However, given the highly customized nature of effective problem-oriented policing interventions, it is important to recognize that the generalizability of specific crime prevention practices identified in an effective application of the approach might be limited, regardless of the evaluation approach used.

The complexity of developing rigorous evaluations in large area evaluations

Another problem noted is that large area studies are not as amenable to large-scale field experiments. If you are working with large treatment areas such as an entire city or neighborhoods suffering from gang violence in a city, how can you develop enough

cases for a valid randomized field trial? For instance, using policing districts as a unit of analysis, there are only five districts with persistent gang homicide problems in Boston (those are B-2, B-3, C-11, D-4, and E-13). Randomly allocating five cases to treatment and control groups would not result in a randomized experiment where stable statistical inferences could be made about the relationship between the treatment and outcomes. This argument does not preclude, of course, strong quasi-experimental designs in which only a small number of matched units are needed. But it does suggest that large experimental field trials are unlikely to be successful when evaluating focused deterrence strategies applied to groups in conflict within larger areal units.

Moreover, group-based focused deterrence strategies intended to reduce citywide levels of gang violence are explicitly designed to deter continued gun violence by gangs not directly subjected to the treatment. These strategies attempt to establish a deterrence regime by diffusing knowledge of enhanced sanction risks associated with specific violent behaviors among a very particular audience. In essence, these focused deterrence strategies attempt to influence the violent behaviors of groups that directly experience treatment and the violent behaviors of groups that vicariously experience treatment through knowledge of what happened to their rivals and allies (Kennedy et al. 1997). In their Campbell review, Braga and Weisburd (2012) noted that the only focused deterrence intervention to investigate the existence of spillover effects on gang violence was the Los Angeles evaluation carried out by Tita et al. (2004). The intervention targeted two rival gangs operating out of the same area (Hollenbeck). Criminal activity (i.e., violent, gang, and gun crimes) was substantially reduced among the two gangs over a 6-month pre-post period. Slightly larger reductions in these crimes were evident among four non-targeted, rival gangs in surrounding areas during the same time period. Part of the explanation for the diffusion effects may rest with fewer feuds between the targeted and non-targeted gangs. The authors also speculated that diffusion effects may have been influenced by social ties among the targeted and rival gangs. This seemed to be especially the case for gang crimes involving guns.

But criticisms of experimental or quasi-experimental methods in this case are much broader, and a number of scholars have begun to argue that experimentation is not likely to yield much benefit. This criticism can be thought of in two contexts. In the first, it raises questions about whether experimental studies can be generalized to other settings, and it is a criticism brought by scholars in many fields (Manski 2013; Sampson et al. 2013; Heckman and Smith 1995). This is the problem of external validity described above. We recognize this difficulty, but it probably has less salience for criticisms regarding more rigorous methods, since all the focused deterrence programs rely on samples that are specific to the jurisdictions of interest. For instance, the nature of the Project Safe Neighborhoods intervention to control gun violence among warring factions of Asian gangs (Braga et al. 2008b) included program elements designed to crack down on the gambling interests of elder gang members that simply were neither present nor needed in Boston's ongoing efforts to control serious violence among its mostly black and Hispanic gangs. Like all studies of problem-oriented crime prevention programs, we simply have to be more careful in generalizing from any studies of programs that are brought to jurisdictions because they have specific characteristics that make them amenable to program development.

A more salient problem in the evaluation of group-based focused deterrence strategies is what some have called the "stable unit treatment value assumption" (SUTVA).

This assumption requires that the treatment or control condition to which a unit is assigned has no impact on the response of another unit (Rubin 1990). This assumption rules out human response to treatments like "Hawthorne" or "John Henry" effects (when participants in the control group alter their behavior purposely as a result of the experiment) and any other forms of social interaction. Berk (2005b) offers several examples of this phenomenon in criminology; for instance, the placement of a substantial number of rival gang members in the same boot camp could dramatically alter the nature of the treatment and the subsequent response. Given that focused deterrence strategies strive to create spillover effects, the inclusion of untreated gangs that were socially connected to treated gangs as comparison groups in an impact evaluation would potentially violate SUTVA.

Our point is that the criticisms that have been raised on the feasibility of particular applications of randomized experiments and stronger quasi-experiments certainly have merit. This is why many of the focused deterrence evaluations used weaker quasi-experimental designs with non-equivalent comparison groups. Given that focused deterrence strategies have only existed for the past 20 years, many valid evaluation questions remain given the nature of the intervention and the criminal behaviors such programs are designed to influence. Reasonable questions include: How can evaluators develop even strong quasi-experimental designs in such complicated contexts? Where can evaluators find equivalent comparison groups? How can evaluators develop a design with enough units for strong experimental comparisons? We focus on these questions by examining how we could improve the rigor of evaluations of specific focused deterrence programs.

Developing more rigorous evaluations of group-based violence reduction programs

A key problem in developing more rigorous evaluations of group-based focused deterrence strategies is the SUVTA assumption we have just reviewed. Can such problems be overcome in these programs? We think one strategy would be simply to exclude potentially "contaminated" gangs from consideration as comparison groups. This is certainly a reasonable way to minimize SUTVA concerns in an evaluation that attempted to determine whether treated gangs changed their violent behaviors relative to the violent behaviors exhibited by untreated gangs. These vicariously-treated gangs, however, offer a much more important opportunity to significantly advance deterrence research as well as program evaluation, and therefore deserve close scrutiny in their own right. The area problem can be solved if we look to use rigorous matching techniques for gangs or gang members.

If the theoretical model underlying group-based focused deterrence is sound, then the program should have two distinct suppression effects on violence: direct suppression through the personal punishment experiences of treated gang members, and indirect suppression through the vicarious punishment experiences of untreated gangs who are socially connected to a treated gang (see Fig. 2). Completed well after the publication of the Campbell focused deterrence review, two companion papers use quasi-experimental methods to unravel these related crime reduction impacts for a reconstituted Operation Ceasefire program in Boston.

As described extensively elsewhere (see, e.g., Braga et al. 2008a), the City of Boston discontinued the Operation Ceasefire strategy in 2000. After several years of rising gang violence, in 2007 the BPD once again implemented the Ceasefire focused deterrence strategy to reduce fatal and non-fatal shootings committed by Boston gangs. Braga et al. (2014) conducted a more rigorous quasi-experimental evaluation of the reconstituted Boston Ceasefire program that used propensity score-matching techniques to develop matched treatment gangs and comparison gangs. Growth-curve regression models with differences-in-differences (DID) estimators were then used to estimate the impact of Ceasefire on gun violence trends for the matched treatment gangs relative to matched comparison gangs during the 2006–2010 study time period. The evaluation reported that total shootings involving directly treated Ceasefire gangs were reduced by a statistically significant 31 % relative to total shootings involving comparison gangs.

The post-2007 version of Boston Ceasefire attempted to create spillover deterrent effects onto other gangs that were socially connected to targeted gangs through rivalries and alliances (Braga et al. 2014). As Ceasefire interventions were completed on targeted gangs, the working group directly communicated to their rivals and allies that "they would be next" if these groups decided to retaliate against treated rival gangs or continue shootings in support of treated allied gangs. These messages were delivered to members of socially-connected gangs via individual meetings with gang members under probation supervision and through direct "street conversations" with gang members by Boston Police officers, probation officers, and gang outreach workers.

In Braga et al. (2014), a propensity score model was used to match Boston street gangs that were the targets of Ceasefire with comparison gangs that were not so targeted. Concerns about SUTVA motivated these authors to exclude all non-Ceasefire gangs that were known to have a rivalry or an alliance with a Ceasefire gang. Because the stated goal of the Braga et al. (2014) evaluation was to estimate the direct effects of Ceasefire on gun violence, and not the program's indirect effects (as depicted in Fig. 2), the inclusion of non-Ceasefire gangs that were socially connected to Ceasefire gangs in the comparison groups would have violated SUTVA. Accordingly, this approach also allows us to avoid the criticism of lack of independence between treatment and control cases.

Braga et al. (2014) used social network analysis data to examine the social connections among the $n=123$ gangs that were involved in at least one shooting in Boston between 2006 and 2010. The post-2007 implementation of Operation Ceasefire directly
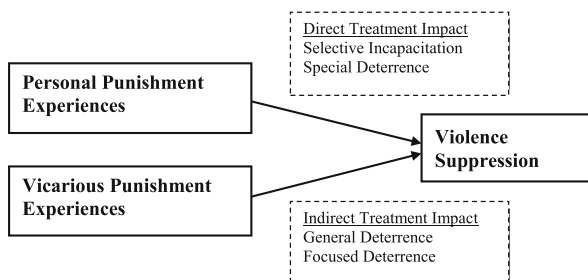
Fig. 2 Conceptual model of the impact of group-based focused deterrence on violence. Adapted from Braga et al. (2013)

applied the focused deterrence strategy to $n=19$ gangs. Twenty-two gangs were socially-connected to the directly-treated Ceasefire gangs through rivalries and alliances. While these socially-connected gangs were not directly subjected to the full Ceasefire treatment, the focused deterrence strategy was designed to reduce their gun violence behaviors via knowledge of what happened to their rivals and allies. As such, these socially-connected gangs can be described as vicariously experiencing Ceasefire treatment. Eighty-two gangs that did not experience direct or vicarious treatment were available to serve as untreated comparison gangs.

Braga et al. (2013) used a similar quasi-experimental design to estimate the spillover deterrence impacts on the gun violence behaviors of the vicariously-treated Boston gangs relative to the gun violence behaviors of untreated Boston gangs. Propensity score-matching routines were used to identify the 19 matched vicariously-treated gangs (86.4 % of 22 socially-connected gangs) and 61 matched comparison gangs (74.4 % of 82 possible comparison groups). Growth-curve regression models with DID estimators were then used to estimate the indirect impact of Ceasefire on gun violence trends for the matched vicariously-treated gangs relative to matched comparison gangs during the same 2006–2010 study time period. The evaluation reported that total shootings involving vicariously-treated Ceasefire gangs were reduced by a statistically significant 24 % relative to total shootings involving comparison gangs.

In some respects, the result from the Braga et al. (2013) study showing an indirect or "spillover effect" of the Ceasefire intervention on gun violence represents a more complete test of deterrence theory than the result showing a direct effect. This is because, as emphasized in Fig. 2, the direct impact of Ceasefire actually comprises two distinct effects: selective incapacitation and special deterrence. Interventions that target violent gang members for prosecution and incarceration can achieve gang-level crime reductions simply by taking the most dangerous and prolific offenders from the targeted gang out of circulation. They can also achieve crime reductions by motivating punished offenders to cease offending or, more likely, to resort to non-violent crimes that draw less attention from law enforcement. However, these two effects are hopelessly confounded, and it appears impossible for any empirical test to untangle them.

From the standpoint of gun violence suppression, of course, the distinction between selective incapacitation and special deterrence is irrelevant; that is, it only matters whether an intervention "works" by increasing public safety (for an argument to this effect, see Miles and Ludwig 2007). From the standpoint of theory, on the other hand, the distinction is of paramount importance. Ceasefire is, after all, touted as a focused *deterrence* intervention as opposed to a selective *incapacitation* intervention. Consequently, because of the empirical ambiguity outlined above, any test of the deterrence efficacy of Ceasefire must, by definition, be evaluated from the spillover effects of the program.

It is also important note here that the companion quasi-experimental evaluations of the post-2007 Ceasefire intervention yielded much more conservative violence reduction estimates when compared to the two-thirds reduction in youth homicides reported in the original 1990s Ceasefire quasi-experimental evaluation (Braga et al. 2001). Consistent with the findings of the Campbell review, the weaker "Level 3" quasi-experimental design was associated with a 51 % larger effect size for the group-based violence reduction strategy ($d=-1.161$; Braga and Weisburd 2012) relative to the effect

sizes reported in the more recent "Level 4" quasi-experimental evaluation of the same type of strategy implemented in the same city ($d=-0.7678$; Braga et al. 2014).

Developing randomized experimental tests of group-based violence reduction programs

The Braga and Weisburd (2012) Campbell review did not identify a single randomized controlled trial of the impacts of a group-based focused deterrence strategy on violence. To our knowledge, this remains true; there have been zero randomized experimental tests of this very well-known and influential approach to violence prevention. Is it feasible to implement randomized experiments in this area?

There seems to be, at least three reasons for the lack of randomized experiments. Two we have already described. First, as described above, the nature of intervention makes it very difficult to randomize the treatment to particular groups without violating SUTVA. Second, these are large area studies, and finding enough treatment and control units will be difficult. But an additional problem is that cities that are willing to adopt group-based violence reduction-focused deterrence strategies are often experiencing high levels of serious gun violence. Implementing a powerful response to gun violence is their primary concern and randomization is generally an afterthought. In some sense, this is related to the generalization problem we noted earlier.

These problems were first identified in the evaluation of the original Boston Ceasefire intervention. As Braga (2013: 232) describes in his account of the Harvard research team's attempt to develop a rigorous evaluation:

> … [we] eventually settled on proposing two versions of a randomized complete block experimental design. In the first proposed design, Boston gangs were matched into pairs based on a variety of group characteristics, such as membership size and violent activity, and then randomly allocated to treatment and control conditions. In the second proposed design, gang turf areas were matched based on a variety of place characteristics, such as area counts of shootings, neighborhood characteristics, and gang size, and then randomly allocated to treatment and control conditions.

The proposal for an evaluation design that included randomly-allocated within-city comparison groups was ultimately rejected over concerns that the Boston Police Department and its partners needed to do everything possible to control all outbreaks of gang violence, and that the intervention, if implemented correctly, would be a powerful deterrent to all gangs in Boston and necessarily contaminate behavior by comparison gangs and in comparison gang turfs.

Given the nature of the intervention and the complexity of varied social connections among gangs than can span multiple areas within a city, multisite cluster randomized trials seem to be a promising approach to conducting more rigorous experimental evaluations of group-based focused deterrence strategies. Multisite experiments are independent randomized controlled trials implemented in two or more sites where evaluators involved in the study plan and collaborate across these sites (Boruch 1997; Weisburd and Taxman 2000). Evaluators

conduct multisite randomized trials to replicate findings from initial single-site studies, to gain a sample size large enough to obtain sufficient statistical power, and to discern moderate to small effect sizes (MacKenzie et al. 2013). Common data collection protocols are used at all sites and one lead center completes the analyses of data from all the sites. Cluster randomized trials have the additional advantage that, if one has enough sites and such sites represent a larger population of sites, significance tests can also be run that can be generalized to that larger population (Weisburd and Taxman 2000). For example, most existing studies make inferences only to the specific site examined, while researchers draw logic models to inferences to the wider population of sites. With a sufficient sample of sites, random effects inferences can be made to the population of sites rather than to the subjects or units under study (Weisburd and Taxman 2000).

Cluster randomized experiments represent a variation of the classic randomized controlled trial design in which clusters (groups) of subjects, rather than individual subjects, are randomly allocated to treatment and control conditions (Murray 1998; Mosteller and Boruch 2002). This design allows better control of treatment "contamination" across individual subjects. In the case of gang violence, this contamination is the SUTVA problem generated by social connections among gangs described above. In a multisite cluster randomized trial, clusters of subjects are randomly allocated to treatment and control conditions in two or more sites. Randomly allocating distinct clusters of gangs connected by rivalries and alliances to treatment and control conditions limits the treatment contamination problem. Researchers in each participating city would need to identify gang conflict and alliance networks and apply social network analysis techniques to specify distinct socially-connected cliques of gangs (see, e.g., Kennedy et al. 1997; Papachristos et al. 2013). Researchers would also need to track shootings by specific gangs during pre-intervention and post-intervention time periods in participating cities.

Since outcomes for gangs within clusters may be correlated, standard sample sizes need to be inflated to for cluster randomized controlled trials. Multisite cluster randomized trials would allow investigators to include much larger numbers of gangs across multiple cities. For instance, a total sample size of 800 gangs (hypothetically, 80 clusters of 10 gangs each) will provide statistical power at the 0.65 level to detect a standardized effect size of 0.20 and statistical power at the 0.99 level to detect a standardized effect size of 0.40, depending on assumptions about the intra class correlations in the outcome measure.[3] With DID estimators, Hierarchical Linear Models (HLMs) can then be used to estimate the difference in fatal and non-fatal gun violence incidents for pre- and post-implementation time periods and between the treatment and control groups (Albright and Marinova 2010; Raudenbush and Bryk 2002).[4]

---

[3] Statistical power estimates were calculated using the "Optimal Design" software available from the University of Michigan (http://sitemaker.umich.edu/group-based/optimal_design_software).
[4] Cluster randomized controlled trials introduce dependence among the subjects within each cluster. In the proposed work, two gangs sampled from the same clique are more likely to be similar in terms of outcomes than gangs sampled from other cliques. HLMs are used to adjust for this lack of independence and account for both individual gang level and gang cluster level covariates.

Developing more rigorous evaluations of focused deterrence programs to reduce repeat offending by high-risk individuals

The two evaluations of focused deterrence strategies designed to reduce offending by high-risk individuals identified by the Campbell review used areas rather than people as the key units of analysis. In Chicago, the PSN treatment was focused on newly-released prisoners with a very high risk of being a victim or offender of gun violence in two treatment police districts (Papachristos et al. 2007). The Chicago PSN evaluation team used propensity score models to match two very similar comparison policing districts on a variety of crime and social indicators. The Newark Ceasefire strategy focused on preventing gun violence by individual gang members in a targeted "Ceasefire Zone" by blending the law enforcement actions with the public health violence prevention activities (Boyle et al. 2010). The comparison zone was identified through spatial analyses of non-fatal gunshot wounds to identify an area of similar size with similar levels of gun violence in Newark and also matched to the Ceasefire Zone based on 2000 census data on the number of block groups in each area, population, resident race and ethnicity, median resident age and household income, concentrated poverty, and vacant housing units.

While these area-level analyses are appropriate to establish program impacts, changes in the violent behaviors are inferred from changes in aggregate outcome measures. Measuring behavioral changes at the individual-level would provide more direct evidence of program impacts. It is important to note here that a recent, unpublished, supplemental quasi-experimental analysis conducted by the Chicago PSN team did examine whether the treatment reduce violent recidivism of program participants (Papachristos et al. 2013). Using survival analyses, the authors found that those who attended a PSN forum were 30 % less likely to be rearrested relative to a comparison group of similar recently released individuals from the same neighborhood. Nevertheless, a randomized controlled trial would be a more rigorous way to estimate individual-level impacts of focused deterrence programs. The Hawaii Opportunity with Probation Enforcement (HOPE) evaluation provides a good example of how to implement and analyze such a randomized experiment (Hawken and Kleiman 2009).

The HOPE intervention was a community supervision program aimed at substance-abusing probationers (Hawken and Kleiman 2009).[5] The program relied on a mandate to abstain from illicit drugs, backed by swift and certain sanctions for drug test failures, and preceded by a clear and direct warning. Probationers were sentenced to drug treatment

---

[5] Based on the Campbell review selection criteria, HOPE was not included in the final Braga and Weisburd (2012) review. However, several scholars contacted during their search for eligible studies believed that HOPE did fit within the general framework of pulling levers-focused deterrence strategies. We agree that it is broadly similar to the Chicago PSN, as both are focused on corrections populations. The key elements of Chicago PSN strategy are administered by the Illinois Department of Correction and the U.S. Attorney's Office (the call-in session is given to returning parolees to selected neighborhoods). The contribution of the Chicago Police Department is limited to increasing their gun policing efforts in the selected neighborhoods. Moreover, probation has a central role in all the gang-/group-based focused deterrence interventions included in our review. Monitoring offenders in the community to ensure they are abiding by probation conditions, changing conditions, and revoking probation are key levers that are pulled in the application of focused deterrence strategies to gangs and criminally active groups. Finally, most applications of pulling levers-focused deterrence strategies have therapeutic elements (e.g., Braga et al. 2001; Papachristos et al. 2007).

only if they continued to test positive for drug use, or if they requested a treatment referral. The deterrence-based HOPE intervention differs significantly from typical drug court operations as it economizes on treatment and court resources. As Hawken and Kleiman (2009) suggest, HOPE does not mandate formal treatment for every proba-tioner, and does not require regularly scheduled meetings with a judge; probationers appear before a judge only when they have violated a rule. HOPE is often linked to the DMI approaches as a related application of focused deterrence (see, e.g., Boyum et al. 2011) as well as gang- and group-based pulling levers-focused deterrence based on the common strategy of certain punishment for offenders (Durlauf and Nagin 2011).

The HOPE evaluation used a randomized controlled trial among general-population substance-abusing probationers where probationers assigned to treatment conditions were compared to probationers assigned to probation-as-usual control conditions (Hawken and Kleiman 2009). The randomized controlled trial used an intent-to-treat design in which all offenders randomly allocated to the treatment condition were included in the HOPE group whether they formally entered the program or not. Of the eligible probationers, two-thirds were assigned to the HOPE treatment ($n=330$) and one-third were assigned to the control group ($n=163$). Ninety-three percent of the probationers assigned for treatment appeared for their initial HOPE warning hearing and participated in the intervention. The experiment commenced in October 2007 and the intervention period lasted for one year.

Based on their analyses of the experimental data, Hawken and Kleiman (2009) concluded that HOPE was very effective in changing the behaviors of substance-abusing probationers. Only 21 % of HOPE probationers experienced new arrests as compared to 47% of control probationers ($p<01$). HOPE proba-tioners outperformed control probationers on a number of other performance measures such as missed probation appointments (treatment=9 %, control= 23 %), positive urine drug test results (treatment=13 %, control=46 %), revo-cation rates (treatment=7 %, control=15 %), and the number of days sentenced to incarceration (treatment=138 days, control=267 days).

A randomized controlled trial of focused deterrence programs intended to change individual violent behaviors could be designed drawing on the key elements of the HOPE evaluation. Indeed, using Chicago PSN as an example, the randomization of offenders into treatment and control groups is very straightforward. Gun- and gang-involved recently released former prison in-mates returning to Chicago neighborhoods could be randomly selected to participate in offender notification meetings. These treatment offenders would then be informed of their vulnerability as felons to federal firearms laws, with stiff mandatory minimum sentences, offered social services, and addressed by community members and ex-offenders.

Developing more rigorous evaluations of drug market intervention focused deterrence programs

The two quasi-experimental evaluations of DMI programs included in the Campbell review used weaker quasi-experimental research designs that compared pre-test and post-test crime outcome trends in treated areas to pre-test and post-test crime outcome

trends in non-equivalent comparison areas. Corsaro and McGarrell (2009) used ARIMA models to analyze crime trends in the following Nashville areas: (1) the McFerrin Park target neighborhood to assess the local effect; (2) adjoining, contagious areas to the McFerrin Park neighborhood to assess whether a local displacement or a diffusion of benefits occurred; and (3) the remainder of Davidson County, once the target and adjoining areas were subtracted from the county totals for general trend comparison purposes. In Rockford (IL), Corsaro et al. (2010) used hierarchical generalized linear growth curve regression models with a dummy variable to represent the implementation of the DMI strategy used to analyze violent and nonviolent crime trends in the treatment Delancey Heights neighborhood relative to violent and nonviolent crime trends in the remainder of Rockford without Delancey Heights.

The seminal DMI-focused deterrence strategy was implemented to control a disorderly and violent drug market operating in the West End neighborhood of High Point, North Carolina. In a simple pre-/post-treatment group-only evaluation, Kennedy and Wong (2009) reported that violent crime decreased 39 % and drug crime decreased by 30 % in the West End. In a paper finished after the Campbell review was completed, Corsaro et al. (2012) significantly advanced the rigor of DMI evaluations by using DID panel regression models to estimate program effects on violent crime trends in treated High Point neighborhoods relative to violent crime trends comparison High Point neighborhoods identified through propensity score matching. In contrast to the large program effect reported by Kennedy and Wong (2009), this more rigorous "Level 4" quasi-experimental evaluation reported a much more modest 12 % reduction in violent crime in the treated areas relative to matched control areas (Corsaro et al. 2012).

The Corsaro et al. (2012) evaluation shows that there are a variety of more rigorous quasi-experimental approaches that researchers can pursue to develop more robust tests of DMI programs.[6] However, it seems relatively straightforward to evaluate DMI programs using place-based randomized controlled trials. Research has demonstrated that urban drug problems are concentrated in very small places. In the Jersey City Drug Market Analysis Project (DMAP), Weisburd and Green (1995) identified 56 drug markets that covered only 4.4 % of the city's street segments but generated about 46 % of both narcotics calls for service and narcotics arrests. These drug markets were randomized in statistical blocks to treatment and control conditions (see Weisburd and Gill 2013). The treatment followed a stepwise approach in which the police sought to engage business owners and residents in drug market disruption activities, applied problem-oriented crackdowns customized to limit local illicit drug selling activities, and maintained subsequent crime control gains by increased patrol attention. Weisburd and Green (1995) found the treatment was associated with significant reductions in disorder calls for service in the treatment drug markets relative to control drug markets.

---

[6] There are, of course, other rigorous quasi-experimental evaluation frameworks that can be applied to place-based policing interventions. For instance, Braga et al. (2011) used propensity score models to match treated high violence street segments and intersections to untreated high violence street segments and intersections in Boston to evaluate a place-based policing program. More recently, Saunders et al. (2014) applied a synthetic control group quasi-experimental design to evaluate the High Point DMI program. The synthetic control approach has been used successfully in political science to measure the economic impact of terrorist conflict in Basque Country (Abadie and Gardeazabal 2003) and tobacco prevention legislation in California (Abadie et al. 2010).

They also noted a diffusion of crime control benefits into two-block catchment areas surrounding the treatment drug markets relative to the control drug markets.

Most place-based randomized controlled trials involve relative small numbers of cases.[7] Randomized controlled trials that use simple randomization schemes face an increased risk, by chance alone, of creating unbalanced treatment and control groups when n is small. Small sample sizes and the increased error generated by unbalanced groups result in experimental designs with low statistical power to detect treatment effects, if in fact they exist (Cohen 1988; Weisburd 1993). Block randomized place-based trials maximize the equivalence of treatment and control groups and improve statistical power (Fisher 1926, 1935). However, in criminological study of places, the fully blocked randomized design is likely to sacrifice degrees of freedom (lost for every restriction on randomization) in cases where data do not allow for a precise subject-to-subject match. Accordingly, Weisburd and Green (1995) used a partially blocked randomized design in their Jersey City DMAP experiment. Weisburd and Gill (2013) used simulation techniques to compare the statistical power associated with randomized block experimental designs relative to randomized trials that use simple randomization schemes. They found that randomized complete block designs resulted in higher powered statistical tests of small n place experiments by creating better balanced treatment and control groups.

More rigorous evaluations of DMI focused deterrence strategies should develop similar randomized block experimental designs used by place-based policing evaluations to test the impact of the DMI intervention on drug market places. Beyond improving estimates of main program effects on crime outcomes, DMI evaluations can also draw upon the approaches developed by place-based policing evaluations to measure crime displacement and diffusion of crime control benefits effects. Place-based policing experiments use DID estimators to analyze pre-test to post-test changes in official crime data in two-block catchment areas immediately surrounding treatment and control crime places (Weisburd and Green 1995; Braga et al. 1999; Braga and Bond 2008). Weisburd et al. (2006) used systematic social observations of drug-selling behaviors in one- and two-block catchment areas surrounding a treated drug market to examine immediate spatial displacement and diffusion effects. They also used ongoing qualitative interviews with drug sellers to gain insights on the mechanisms through which offenders might be deterred or discouraged from continuing their drug selling behaviors.

While it is laudable that the Nashville and High Point studies measured potential displacement and diffusion effects, the current state-of-the-art in DMI evaluations is not nearly as sophisticated. The Nashville evaluation examined possible displacement and diffusion effects via time series analysis models to estimate pre-/post-intervention changes in crime outcome trends in a large adjoining area [comprised eight police "sub-beats" that covered 3.6 square miles (9.2 sq km); Corsaro and McGarrell 2009]. The High Point evaluation used a DID estimator to compare violent crime trends in census blocks immediately contiguous to targeted areas throughout the city relative to violent crime trends in remainder of the city (Corsaro et al. 2012). The High Point found no evidence of significant crime displacement. The Nashville evaluation reported

---

[7] For instance, the Jersey City problem-oriented policing in violent places (Braga et al. 1999) and Lowell policing crime and disorder hot spots (Braga and Bond 2008) randomized experiments involved only 24 and 35 places, respectively.

statistically significant reductions in illegal drug possession offenses, drug equipment offenses, and total calls for service in the adjoining area. This suggested that the Nashville DMI intervention was associated with a noteworthy diffusion of crime control benefits beyond the McFerrin Park target neighborhood.

Conclusion

This paper has provided a critical assessment of the existing body of focused deterrence evaluation research and has suggested analytical approaches that could be used to develop more rigorous controlled evaluations. We identified three general issues that have been implicated in the proliferation of weaker quasi-experimental evaluations of focused deterrence programs. These include developing appropriate units of analysis, generalizing findings beyond the study site, and SUTVA concerns. Focused deterrence programs designed to reduce repeat offending by high-rate offenders and control drug market hot spot locations easily lend themselves to randomized controlled trial designs. Group-based violence reduction strategies, however, are much more complex to evaluate using more rigorous designs due to SUTVA concerns. These types of focused deterrence strategies tend to be implemented citywide and designed to maximize deterrence spillover effects to untreated gangs and criminally active groups. Counterfactual models based on propensity score analysis and multisite cluster randomized controlled trials offer very promising methodological approaches that could be used to good effect in providing stronger estimates of focused deterrence program impacts.

It is worth noting here that the quality of quasi-experimental evaluations of focused deterrence strategies have improved greatly over time. Contemporary quasi-experimental evaluations of focused deterrence strategies use sophisticated statistical matching techniques, panel designs, and higher-powered statistical models (e.g., Corsaro et al. 2012; Braga et al. 2011, 2013). Existing focused deterrence evaluations, however, are generally ex-post facto assessments of programs that were implemented with little a priori thought by implementers given to conducting rigorous evaluations. Journal articles and reports on the quasi-experimental evaluations examined here uniformly describe cities struggling with very serious crime problems. Policy makers and practitioners in these cities respond to these crises by implementing programs that they believe have the best chance of addressing their chronic crime problem in the near term. Program evaluation is, at best, an afterthought. As such, in the past, focused deterrence evaluators were forced to do the best job they could with the situation they inherited. None of the evaluators described an experience that would lend itself to the considerable upfront planning that randomized field experiments require.

We believe that it is possible for cities to implement focused deterrence strategies in such way that harm can be reduced in the near term and a randomized field experiment can be implemented. As described earlier, focused deterrence strategies are usually customized to local conditions through upfront problem analysis. While data are being collected and analyzed for the problem analysis, randomized field experiments can be designed and ready to implement. Indeed, the randomized experiments described here can serve as blueprints for future research and development efforts. The remaining obstacles to implementing randomized field experiments are often political and ethical. Politicians and police executives might be wary of community backlash to any decision

that withholds a potentially effective program for all communities that need it. They also might be uncomfortable with the idea that people could die and communities be further harmed if they are unlucky enough to be in the control group.

It is equally problematic, however, to implement programs that falsely raise citizen expectations of large violent crime reductions and dramatic changes in the quality of residential life in neighborhoods suffering from persistent drug and violent crime problems. As Phil Cook (2012: 162), "the quest for a miracle cure for crime and violence sometimes leads to an early or excessive embrace of an unproven technology." It is much more prudent to take a skeptical approach to policy interventions until a portfolio of proven practices has been developed. The available DMI evaluation evidence suggests that the approach does indeed reduce crime. However, it remains unclear whether the DMI generates the large violence reduction and quality of life improvements described by Kennedy (2008) or if the impacts are much smaller, as documented by Corsaro et al. (2012).

The federal government could play a powerful role in supporting randomized controlled trials of focused deterrence strategies. Unfortunately, to date, using federal funds to advance the possibility of a randomized field experiment has not produced the desired effects. For instance, the U.S. National Institute of Justice is supporting a comprehensive evaluation of the DMI focused deterrence program in 12 sites, each of which received technical assistance in the SMI from the U.S. Bureau of Justice Assistance via the School of Criminal Justice at Michigan State University.[8] The Rand Corporation serves as the evaluator of the NIJ-supported DMI initiative and, in their proposal, advocated the use of "randomization across and within jurisdictions will maximize our potential to draw strong inferences and make sound policy recommendations."[9] Regrettably, not a single participating site was willing to participate in a randomization scheme either across cities or within cities.[10] Future solicitations for focused deterrence demonstration programs should make participation in a randomized field experiment a necessary condition to receive federal funds.

It is helpful to consider the trajectory of closed-circuit television (CCTV) evaluations when lamenting the absence of randomized experiments from the available evaluation evidence on focused deterrence programs. In their systematic review of the effectiveness of CCTV on crime, Welsh and Farrington (2009) did not identify a single randomized controlled trial and relied exclusively on quasi-experimental evaluations to draw conclusions about the crime prevention efficacy of these programs. Further, they expressed concern that a majority of the quasi-experiments used weak "Level 3" designs. However, a few years after the release of the Welsh and Farrington (2009) updated systematic review, the Urban Institute conducted the very first randomized controlled trial of the impact of CCTV on crime (LaVigne and Lowry 2011). We hope that our call inspires randomized experimentation in this important area.

Our recommendations suggest, more broadly, that evaluators do not need to "settle for less" when developing empirical tests of large area-based crime prevention programs.

---

[8] http://www.dmimsu.com/
[9] http://grants.ojp.usdoj.gov:85/selector/awardDetail?awardNumber=2010-DJ-BX-1672&fiscalYear=2010&applicationNumber=2010-94093-CA-IJ&programOffice=NIJ&po=NIJ
[10] Personal communication with Jessica Saunders of the Rand Corporation (October 19, 2013).

More rigorous evaluations can, and should, be developed and implemented. Many of our observations in this essay can be applied to other crime and justice interventions, such as problem-oriented policing, where the existing body of evidence is characterized by a preponderance of weak evaluation designs (Weisburd et al. 2008). Weak evaluations, unfortunately, provide less valid answers to policy questions when compared to well-designed quasi-experiments and randomized controlled trials (Shadish et al. 2002). A number of crime and justice scholars suggest that there is a "moral imperative" in pursuing the most rigorous evaluation designs to discover whether a program is effective (see, e.g., Boruch 1975; Weisburd 2003). Moreover, as noted by Joan McCord (2003), unproven programs can sometimes produce harmful effects, and rigorous evaluations, most notably randomized experiments, are necessary to identify any beneficial or harmful effects. Isolating the effects of treatments or programs from other confounding aspects of selection or design is viewed as one of the evaluator's most important obligations to society. When the evaluation evidence base is largely informed by weak designs, practitioners risk implementing certain treatments or programs as effective crime prevention practices when they are not; this can lead to significant economic and social costs.

References

Abadie, A., & Gardeazabal, J. (2003). The economic costs of conflict: a case study of the Basque country. *American Economic Review, 93*, 113–132.

Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. *Journal of the American Statistical Association, 105*, 493–505.

Albright, J., & Marinova, D. (2010). *Estimating multilevel models using SPSS, Stata, SAS, and R*. http://www.indiana.edu/~statmath/stat/all/hlm/hlm.pdf.

Berk, R. (2005a). Knowing when to fold 'em: an essay on evaluating the impact of ceasefire, compstat, and exile. *Criminology & Public Policy, 4*, 451–466.

Berk, R. A. (2005b). Randomized experiments as the bronze standard. *Journal of Experimental Criminology, 1*, 417–433.

Boruch, R. F. (1975). On common contentions about randomized field experiments. In R. F. Boruch & H. L. Reicken (Eds.), *Experimental testing of public policy: The proceedings of the 1974 social sciences research council conference on social experimentation* (pp. 107–142). Boulder, CO: Westview Press.

Boruch, R. F. (1997). *Randomized experiments for planning and evaluation*. Newbury Park, CA: Sage.

Boyle, D. J., Lanterman, J., Pascarella, J., & Cheng, C. C. (2010). *The impact of Newark's operation ceasefire on trauma center gunshot wound admissions*. Newark, NJ: University of Medicine and Dentistry of New Jersey, Violence Institute of New Jersey.

Boyum, D. A., Caulkins, J. P., & Kleiman, M. (2011). Drugs, crime, and public policy. In J. Q. Wilson & J. Petersilia (Eds.), *Crime and public policy* (pp. 368–410). New York: Oxford University Press.

Braga, A. A. (2008). Pulling levers focused deterrence strategies and the prevention of gun homicide. *Journal of Criminal Justice, 36*, 332–343.

Braga, A. A. (2010). Setting a higher standard for the evaluation of problem-oriented policing initiatives. *Criminology & Public Policy, 9*, 173–182.

Braga, A. A. (2012). Getting deterrence right? evaluation evidence and complementary crime control mechanisms. *Criminology & Public Policy, 11*, 201–210.

Braga, A. A. (2013). Quasi-experimentation when random assignment is not possible: observations from practical experiences in the field. In B. C. Welsh, A. A. Braga, & G. Bruinsma (Eds.), *Experimental criminology: prospects for improving science and public policy* (pp. 223–252). New York: Cambridge University Press.

Braga, A. A., & Bond, B. J. (2008). Policing crime and disorder hot spots: a randomized controlled trial. *Criminology, 46*, 577–607.

Braga, A. A., & Weisburd, D. (2012). *The effects of "pulling levers" focused deterrence strategies on crime. Campbell Systematic Reviews.* doi:10.4073/csr.2012.6.

Braga, A. A., Weisburd, D. L., Waring, E. J., Green-Mazerolle, L., Spelman, W., & Gajewski, F. (1999). Problem-oriented policing in violent crime places: a randomized controlled experiment. *Criminology, 37*, 541–580.

Braga, A. A., Kennedy, D. M., Waring, E. J., & Piehl, A. M. (2001). Problem-oriented policing, deterrence, and youth violence: an evaluation of Boston's operation ceasefire. *Journal of Research in Crime and Delinquency, 38*, 195–225.

Braga, A. A., Hureau, D. M., & Winship, C. (2008a). Losing faith? police, black churches, and the resurgence of youth violence in Boston. *Ohio State Journal of Criminal Law, 6*, 141–172.

Braga, A. A., Pierce, G., McDevitt, J., Bond, B., & Cronin, S. (2008b). The strategic prevention of gun violence among gang-involved offenders. *Justice Quarterly, 25*, 132–162.

Braga, A. A., Hureau, D. M., & Papachristos, A. V. (2011). An ex-post-facto evaluation framework for place-based police interventions. *Evaluation Review, 35*, 592–626.

Braga, A. A., Apel, R., & Welsh, B. (2013). The spillover effects of focused deterrence on gang violence. *Evaluation Review, 37*, 314–342.

Braga, A. A., Hureau, D. M., & Papachristos, A. V. (2014). Deterring gang-involved gun violence: measuring the impact of Boston's operation ceasefire on street gang behavior. *Journal of Quantitative Criminology, 30*, 113–139.

Campbell, D. T., & Boruch, R. F. (1975). Making the case for randomized assignment to treatment by considering the alternatives. In C. Bennett & A. Lumsdaine (Eds.), *Evaluation and experiments: some critical issues in assessing social programs* (pp. 195–296). New York: Academic.

Clarke, R. V. (Ed.). (1997). *Situational crime prevention: successful case studies*. New York: Harrow and Heston.

Clarke, R. V., & Cornish, D. (1972). *The controlled trial in institutional research*. London: H.M. Stationary Office.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cook, P. J. (2012). Editorial introduction: the impact of drug market pulling levers policing on neighborhood violence. *Criminology & Public Policy, 11*, 161–164.

Cook, T. D., & Campbell, D. T. (1979). *Quasi-experimentation: design and analysis issues for field settings*. Chicago: Rand McNally.

Cook, P. J., & Ludwig, J. (2006). Aiming for evidence-based gun policy. *Journal of Policy Analysis and Management, 48*, 691–735.

Corsaro, N., & McGarrell, E. (2009). *An evaluation of the Nashville drug market initiative (DMI) pulling levers strategy.* East Lansing, MI: Michigan State University.

Corsaro, N., Brunson, R., & McGarrell, E. (2010). Problem-oriented policing and open-air drug markets: examining the Rockford pulling levers strategy. *Crime & Delinquency.* doi:10.1177/0011128709345955.

Corsaro, N., Hunt, E., Hipple, N. K., & McGarrell, E. (2012). The impact of drug market pulling levers policing on neighborhood violence: an evaluation of the high point drug market intervention. *Criminology & Public Policy, 11*, 167–200.

Durlauf, S., & Nagin, D. (2011). Imprisonment and crime: can both be reduced? *Criminology & Public Policy, 10*, 13–54.

Eck, J. (2002). Learning from experience in problem-oriented policing and situational prevention: the positive functions of weak evaluations and the negative functions of strong ones. In N. Tilley (Ed.), *Evaluation for crime prevention, crime prevention studies* (Vol. 14, pp. 93–117). Monsey, NY: Criminal Justice Press.

Engel, R. S., Corsaro, N., & Skubak Tillyer, M. (2010). *Evaluation of the Cincinnati initiative to reduce violence (CIRV)*. Cincinnati, OH: University of Cincinnati Policing Institute.

Fagan, J. (2002). Policing guns and youth violence. *The Future of Children, 12*, 133–151.

Farrington, D. P., Gottfredson, D. C., Sherman, L. W., & Welsh, B. C. (2006). The Maryland scientific methods scale. In L. W. Sherman, D. P. Farrington, B. C. Welsh, & D. L. MacKenzie (Eds.), *Evidence-based crime prevention* (revth ed., pp. 13–21). New York: Routledge.

Fisher, R. A. (1926). The arrangement of field experiments. *Journal of the Ministry of Agriculture of Great Britain, 33*, 503–513.

Fisher, R. A. (1935). *The design of experiments*. Edinburgh: Oliver and Boyd.

Goldstein, H. (1990). *Problem-oriented policing*. Philadelphia, PA: Temple University Press.

Guerette, R. T. (2009). The pull, push, and expansion of situational crime prevention evaluation: an appraisal of thirty-seven years of research. In J. Knutsson & N. Tilley (Eds.), *Evaluating crime reduction initiatives, crime prevention studies* (Vol. 24, pp. 29–58). Monsey, NY: Criminal Justice Press.

Harless, W. (2013). *Cities use sticks, carrots to rein in gangs*. October: Wall Street Journal. 14.

Hawken, A. & Kleiman, M. (2009). *Managing drug involved probationers with swift and certain sanctions*. Final report submitted to the National Institute of Justice. Unpublished report.

Heckman, J., & Smith, J. (1995). Assessing the case for social experiments. *Journal of Economic Perspectives, 9*, 85–110.

Kennedy, D. (1997). Pulling levers: chronic offenders, high-crime settings, and a theory of prevention. *Valparaiso University Law Review, 31*, 449–484.

Kennedy, D. (2008). *Deterrence and crime prevention*. New York: Routledge.

Kennedy, D., & Wong, S.-L. (2009). *The high point drug market intervention strategy*. Washington, DC: Community Oriented Policing Services, U.S. Department of Justice.

Kennedy, D., Piehl, A., & Braga, A. A. (1996). Youth violence in Boston: gun markets, serious youth offenders, and a use-reduction strategy. *Law & Contemporary Problems, 59*, 147–196.

Kennedy, D. M., Braga, A. A., & Piehl, A. M. (1997). The (un)known universe: mapping gangs and gang violence in Boston. In D. L. Weisburd & J. T. McEwen (Eds.), *Crime mapping and crime prevention* (pp. 219–262). Monsey, NY: Criminal Justice Press.

Knutsson, J. (2009). Standards of evaluations in problem-oriented policing projects: Good enough? In J. Knutsson & N. Tilley (Eds.), *Evaluating crime reduction initiatives, Crime prevention studies* (24th ed., pp. 7–28). Monsey, NY: Criminal Justice Press.

LaVigne, N., & Lowry, S. (2011). *Evaluation of camera use to prevent crime in commuter parking facilities: a randomized controlled trial*. Washington, DC: Urban Institute.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Ludwig, J. (2005). Better gun enforcement, less crime. *Criminology & Public Policy, 4*, 677–716.

MacKenzie, D. L., Umamaaheswar, J., & Lin, L.-C. (2013). Multisite randomized trials in criminology. In B. C. Welsh, A. A. Braga, & G. Bruinsma (Eds.), *Experimental criminology: prospects for improving science and public policy* (pp. 163–193). New York: Cambridge University Press.

Manski, C. F. (2013). *Public policy in an uncertain world: analysis and decisions*. Cambridge, MA: Harvard University Press.

McCord, J. (2003). Cures that harm: unanticipated outcomes of crime prevention programs. *Annals of the American Academy of Political and Social Science, 587*, 16–30.

McGarrell, E., Chermak, S., Wilson, J., & Corsaro, N. (2006). Reducing homicide through a 'lever-pulling' strategy. *Justice Quarterly, 23*, 214–229.

Miles, T., & Ludwig, J. (2007). The silence of the lambdas: deterring incapacitation research. *Journal of Quantitative Criminology, 23*, 287–301.

Morgan, S. L., & Winship, C. (2007). *Counterfactuals and causal inference: methods and principals for social research*. New York: Cambridge University Press.

Mosteller, F., & Boruch, R. F. (2002). *Evidence matters: randomized trials in educationresearch*. Washington, DC: Brookings Institution.

Murray, D. M. (1998). *Design and analysis of group-randomized trials*. New York: Oxford University Press.

Papachristos, A. V., Meares, T., & Fagan, J. (2007). Attention felons: evaluating project safe neighborhoods in Chicago. *Journal of Empirical Legal Studies, 4*, 223–272.

Papachristos, A. V., Wallace, D., Meares, T., & Fagan, J. (2013). Desistance and legitimacy: The impact of offender notification meetings on recidivism among high risk offenders. Unpublished manuscript.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London: Sage.

Raudenbush, S. W., & Bryk, T. (2002). *Hierarchical linear models: applications and data analysis methods* (2nd ed.). Newbury Park, CA: Sage.

Rosenbaum, P., & Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika, 70*, 41–55.

Rosenbaum, P., & Rubin, D. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *American Statistician, 39*, 33–38.

Rosenfeld, R., Fornango, R., & Baumer, E. (2005). Did ceasefire, compstat, and exile reduce homicide? *Criminology & Public Policy, 4*, 419–450.

Rubin, D. B. (1990). Formal modes of statistical inferences for causal effects. *Journal of Statistical Planning Inference, 25*, 279–292.

Sampson, R. J. (2010). Gold standard myths: observations on the experimental turn in quantitative criminology. *Journal of Quantitative Criminology, 26*, 489–500.

Sampson, R. J., Winship, C., & Knight, C. (2013). Translating causal claims: principles and strategies for policy-relevant criminology. *Criminology & Public Policy, 12*, 587–616.

Saunders, J., Lundberg, R., Braga, A. A., Ridgeway, G., & Miles, J. (2014). *A synthetic control approach to evaluating multiple geographically-focused crime interventions in the same city: DMI in high point*. Santa Monica, CA: Rand Corporation.

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for general causal inference*. Boston: Houghton-Mifflin.

Sherman, L. W., Gottfredson, D. C., MacKenzie, D. L., Eck, J. E., Reuter, P., & Bushway, S. D. (1997). *Preventing crime: what works, what doesn't, what's promising*. Washington, DC: U.S. Department of Justice, National Institute of Justice.

Skogan, W., & Frydl, K. (2004). *Fairness and effectiveness in policing: the evidence committee to review research on police policy and practices*. Washington, DC: The National Academies Press.

Tilley, N. (2009). What's the "what" in "what works?" Health, policing, and crime prevention. In J. Knutsson & N. Tilley (Eds.), *Evaluating crime reduction initiatives, Crime prevention studies* (24th ed., pp. 121–146). Monsey, NY: Criminal Justice Press.

Tita, G., Riley, K. J., Ridgeway, G., Grammich, C., Abrahamse, A., & Greenwood, P. W. (2004). *Reducing gun violence: results from an intervention in east Los Angeles*. Santa Monica: RAND Corporation.

Weisburd, D. (1993). Design sensitivity in criminal justice experiments. In M. Tonry (Ed.), *Crime and justice: a review of research* (Vol. 17, pp. 337–379). Chicago: University of Chicago Press.

Weisburd, D. (2003). Ethical practice and evaluation of interventions in crime and justice: the moral imperative for randomized trials. *Evaluation Review, 27*, 336–354.

Weisburd, D. (2010). Justifying the use of non-experimental methods and disqualifying the use of randomized controlled trials: challenging the folklore in evaluation research in crime and justice. *Journal of Experimental Criminology, 6*, 209–27.

Weisburd, D., & Eck, J. (2004). What can police do to reduce crime, disorder and fear? *Annals of the American Academy of Political and Social Science, 593*, 42–65.

Weisburd, D., & Gill, C. (2013). Block randomized trials at places: rethinking the limitations of small N experiments. *Journal of Quantitative Criminology*. doi:10.1007/s10940-013-9196-z.

Weisburd, D., & Green, L. (1995). Policing drug hot spots: The Jersey City drug market analysis experiment. *Justice Quarterly, 12*, 711–735.

Weisburd, D., & Taxman, F. (2000). Developing a multi-center randomized trial in criminology: the case of HIDTA. *Journal of Quantitative Criminology, 16*, 315–339.

Weisburd, D., Lum, C. M., & Petrosino, P. (2001). Does research design affect study outcomes in criminal justice? *Annals of the American Academy of Political and Social Science, 578*, 50–70.

Weisburd, D., Wyckoff, L., Ready, J., Eck, J. E., Hinkle, J. C., & Gajewski, F. (2006). Does crime just move around the corner? a controlled study of spatial displacement and diffusion of crime control benefits. *Criminology, 44*, 549–592.

Weisburd, D., Telep, C., Hinkle, J., & Eck, J. (2008). *The effects of problem-oriented policing on crime and disorder. Campbell Systematic Reviews*. doi:10.4073/csr.2008.14.

Wellford, C. F., Pepper, J. V., & Petrie, C. V. (Eds.). (2005). *Firearms and violence: a critical review. committee to improve research information and data on firearms*. Washington, DC: The National Academies Press.

Welsh, B. C., & Farrington, D. P. (2009). *Making public places safer: surveillance and crime prevention*. New York: Oxford University Press.

Welsh, B. C., Peel, M. E., Farrington, D. P., Elffers, H., & Braga, A. A. (2011). Research design influence on study outcomes in crime and justice: a partial replication with public area surveillance. *Journal of Experimental Criminology, 7*, 183–198.

Wilkinson, L., & Task Force on Statistical Inference. (1999). Statistical methods in psychology journals: guidelines and expectations. *American Psychologist, 54*, 594–604.

Anthony A. Braga is the Don M. Gottfredson Professor of Evidence-Based Criminology in the School of Criminal Justice at Rutgers University and a Senior Research Fellow in the Program in Criminal Justice Policy and Management at Harvard University.

David L. Weisburd is the Walter E. Meyer Professor of Law and Criminal Justice at Hebrew University Law School and Distinguished Professor in the Criminology, Law, and Society Department at George Mason University.